Assignment V : Support Vector Machines

March 27, 2025

ID5002W: Industrial AI Laboratory

Total Marks: 80

Instructions

1. Assignment shall be submitted before the due date. Late submissions will not be entertained. If you cannot submit the assignment due to some reasons, please contact Dr. Tirthankar by email.

2. All the assignments must be the student's own work. The students are encouraged to discuss or consult friends or classmates. However, they have to submit their own work. Any malpractice will be reported to the authorities and action will be taken as per the IIT Madras rules.

3. If you find the solution in the book or article or on the website, please indicate the reference in the solution

4. You are expected to submit "*.py" file instead of notebook. Follow the steps mentioned below:

    (a) Use the datasets provided to you as per your roll number for solving the assignment problems on regression and classification.

    (b) Create one notebook for Problem 1 and another notebook for Problem 2.

    (c) If you have used

    ```
    df = pd.read_csv("brain_tumor_dataset.csv")
    ```

    where "brain_tumor_dataset.csv" is the file provided to you, replace the code with the following before submitting the assignment.
    For Problem 1,

    ```
    dataset ="./data/problem1/brain_tumor_dataset.csv"
    df = pd.read_csv(dataset)
    ```

    For Problem 2,

    ```
    dataset ="./data/problem2/brain_tumor_dataset.csv"
    df = pd.read_csv(dataset)
    ```

    (d) Now, download each notebook as a "*.py" file and submit both the "*.py" files.

5. A short report in pdf format along with the code should be submitted containing results and analysis as asked in the assignment.

6. Please note:

   - Code should execute without any error.
   - The code should be clean with readable comments.
   - The output of the code should be in a relevant format so that it can be understood by an evaluator.
   - *The path mentioned in item 4(c) is case sensitive and should be correctly written in the code. Also, the submissions should be "*.py" files. If the path is incorrect or the submissions are not "*.py" files, then marks will be deducted from those allotted for code correctness, as deemed fit.*
   - For figures, **do not** use plt.show(). Instead, use an appropriate command to save the figure, such as:
     plt.savefig("roll_no_hyperplane.png") or plt.savefig("roll_no_hyperplane.pdf").
     The evaluator should understand the figure content from the file name. The entire string should be in lower case.

7. Grading Policy:

   - 50% code correctness.
   - 10% code readability and comments.
   - 40% on report and analysis.

Problem 1

The dataset for the assignment is a brain tumor data (brain_tumor_dataset.csv) that has 20 columns and 20000 rows. The first objective is to classify the data based on tumor type. You can use the 80:20 train-test split ratio.

(a) Build a Support vector machine classifier model to classify whether a tumor is benign or malignant. [**Marks:15**]

(b) Plot the AUC-ROC curve for the model built. Also print the classification report. [**Marks:10**]

(c) Use PCA to reduce the dimension to 2 features. Build a classifier model with this data and plot the hyperplane and margins. [**Marks:15**]

Problem 2

The second objective is to predict the survival rate based on the tumor data provided.

(a) Use a Support Vector Regressor model to predict the survival rate. [**Marks: 15**]

(b) Report the accuracy and R^2 score of the model. [**Marks: 10**]

(c) Compare and plot the accuracy of the model built with that of a linear regression model. [**Marks: 15**]